TECHNOLOGY BRIEF

Data Center Memory and Storage
Intel® Optane™ Technology

Faster Access to More Data

Intel® Optane™ technology drastically reduces latency to data at the memory, device, and computing system level.

Frank T. Hady, Ph.D.
Intel Fellow
Chief Optane Systems Architect

Intel Non-Volatile Memory Solutions Group

From the very beginning, computing systems have exploited the tendency of algorithms to access some data more than other data (temporal locality), and to access data stored close together (spatial locality), to increase performance and reduce cost. Data accessed more often is held in faster, smaller memories, while bulk data accessed less frequently is held in storage, costing much less per bit. This interplay between the memories within a computing system—and software running on that system—is commonly called the memory and storage hierarchy.

At every level in this hierarchy, both bandwidth and latency of data access matter. Over the years, sustained improvements in existing memory and storage technologies have produced continual increases in bandwidth, while latencies have remained much more constant. With the introduction of Intel® Optane™ technology as both memory and storage, this hierarchy is seeing its most significant addition since the introduction of NAND SSDs – or arguably since the introduction of DRAM itself. This paper explores the historically low latency offered by Intel® Optane™ SSDs first, and then delves into the system level advantages of Intel® Optane™ DC persistent memory—the same underlying Intel® Optane™ media deployed as system memory.

Memory and Storage Technical Series

The Direct Connection to Intel Fellows and Principal Engineers

This paper is part of a series designed to help system architects, engineers, and IT administrators understand the technological limitations of traditional memory and storage, how those limitations have led to performance and capacity gaps in the data center, and how Intel® Optane™ technology helps fill those gaps with a new industry-disrupting architecture.

The Memory and Storage series examines several topics that affect memory and storage performance and capacity, including bandwidth, latency, queue depth, quality of service (QoS), and reliability.

Intel® Optane™ DC SSDs Significantly Reduce Storage Latency

Intel Optane technology delivers unprecedented reductions in storage latency, reductions not just in the average latency an application sees but also (even more) in the infrequent longer latencies that occur under heavy system loads.

Figure 1 shows a comparison of two SSDs, a leading NAND-based Intel® SSD DC P4610 and an Intel® Optane™ SSD DC P4800X.[1] This complex latency vs. load chart requires some explanation.

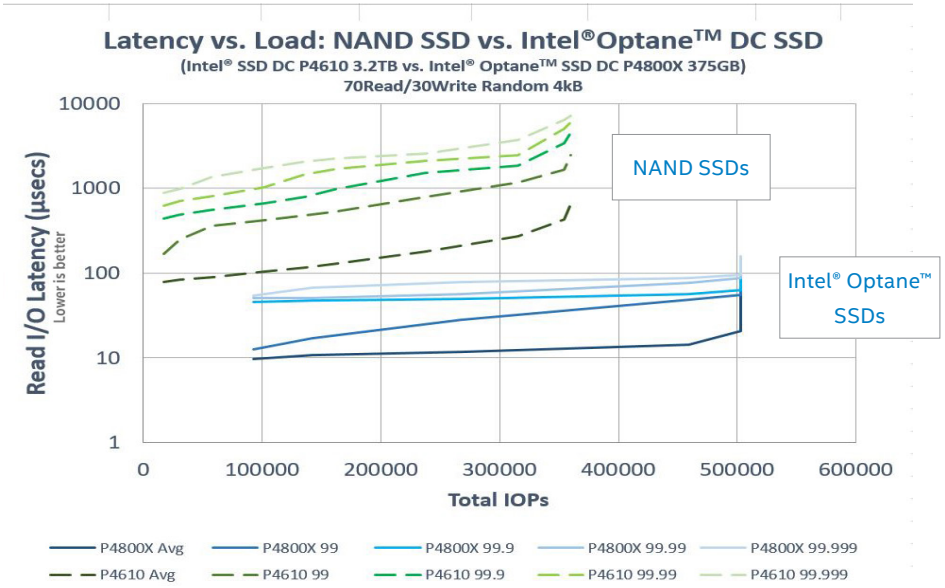


Figure 1. Intel Optane SSDs deliver much lower latency at load than NAND-based SSDs[1]

Read Latency Most Often Determines System Performance

The x-axis is the total mixed read/write IOPs delivered by the SSD under test. For each load (IOP level delivered) the latency of the read I/Os is measured using a flexible I/O tool (FIO), to data return, and plotted using the y-axis. We measure read latencies because software often needs the data requested to make progress, that's why it reads the data in the first place. Writes, on the other hand, are quickly guaranteed persistent by most SSDs, which hold them in a SRAM, assured to be flushed to non-volative memory, even in the event of a power failure. So it is read latency that most often determines system performance.

The performance for each SSD is plotted as a family of lines (solid blue for Intel Optane SSD and dashed green for NAND SSD). The lines represent average latency measured and the longer, less frequently seen read latencies (often called tails since they appear that way in the right side of a histogram of latencies). The time to complete a read depends on many factors (the memory technology, how busy the SSDs are, etc.) and so varies. It's important to understand average latency and also the latency distribution. For each SSD, the bottom line is the average—or 50th percentile—latency.

The next line up is the 99th percentile latency, and so on up to the top line representing 99.999th percentile latency. 99.999th percentile latency corresponds to the next-to-the-slowest read out of 100,000 reads—an infrequent event that matters to some applications.

Finally, don't miss the labeling on the y-axis, a semi-log plot so each grid line represents a 10x difference from the previous. Now that Figure 1 is understood, we can use it to draw a few important conclusions.

Intel® Optane™ Memory Media Provides Faster Access

Intel® Optane™ memory media delivers data significantly faster than NAND media. And the Intel Optane SSD has been designed to provide that advantage to the rest of the computing system through a hardware-only SSD read/write path through the SSD controller, unlike the firmware-involved path found in NAND SSD controllers.

The average latency line for each SSD (the bottom lines in Figure 1) shows that the Intel Optane SSD delivers greater than 10x latency advantage (measured using a flexible I/O tool, or FIO) at every 70/30 read/write IOPs level shown. This means data is returned, on average in just 1/10th the time. Intel Optane memory technology is byte addressable and allows rapid, in-place writes. NAND requires large block reads and a very large block erases before relatively slow writes. Data returned faster means less time waiting and faster application execution time.

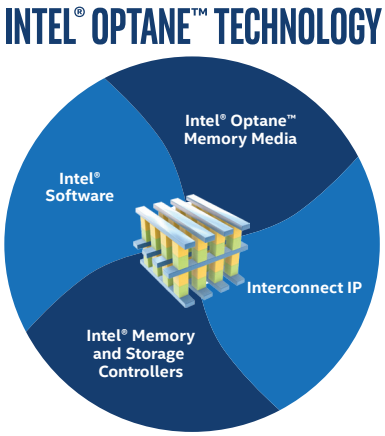
Specification sheets for SSDs often quote nominal or idle latency. Figure 1 shows this latency as the left most point of the average latency line, this is the queue depth equals one (QD=1) point in the curve. When measuring this point, FIO asked for a single 4 KB read or write, and waitied until that transfer completed before asking for another, using just a single thread. Notice for QD=1, the Intel Optane SSD delivers far higher IOPs than the NAND SSD. This is precisely because it offers lower latency. In fact, this is the impact of that lower latency to throughput. This also means that QD=1, or really any constant QD, is hardly a fair comparison point, since the Optane SSD is completing far more work (more IOPS) than the NAND SSD. SSD latencies should be compared under the same load like Figure 1.

Using this media advantage, Intel Optane SSDs avoid long waits for writes to complete and the garbage collection inherent in NAND SSDs. This means that Intel Optane SSDs don't face the background writes, and the delay to reads, seen by NAND SSDs. These advantages translate to significant quality of service (QoS) advantages described above. QoS is visible in the charts through the height and shape of the family of lines extending above the average latency line for each SSD. The NAND SSD has a 99.999th percentile latency that is more than 10x the average latency, while the Intel Optane SSD latency distribution stays within that 10x threshold. More importantly the Intel Optane SSD lines, at all percentiles, are much flatter than the NAND SSD. This more ideal curve means that the Intel Optane SSD delivers steadier performance, even under heavy loads.

Intel® Optane™ Technology Breaks Through the NAND Barrier

Intel Optane technology is built on a completely new memory media that supports write in place, is byte addressable like DRAM, non-volatile like NAND, and has a read/write latency between the two.

Intel Optane technology combines Intel Optane memory media with Intel® controllers, software, and system interconnects that can be deployed as memory or storage.



For many applications, lower SSD latency, and better SSD QoS translates to better performance. For some applications, we see the lower average latency result in decreased runtimes, with less total time spent waiting for storage access to return. In this case, the Intel Optane SSD delivers a better user experience. For other applications where user responsiveness requirements exist, and where each user visible operation requires multiple accesses, we see the improved QoS result in an ability to support many more users before the responsiveness requirements are exceeded. In this case the Intel Optane SSD allows the system to support more users for a cost advantage.

Software Latency: The New Frontier

When accessing SSD resident data—even for modern Intel Optane DC SSDs—there are several software layers in the operating system, between the drive and the application, that contribute to latency (see Figure 2). This software is commonly called the "storage stack".

When an application initiates a read or write, that request is handed off to the operating system. Together, the instructions executed in the storage stack can take a total of up to 4–10 microseconds (µs), or even longer, of CPU work. This time can be as large or larger than the time required to move data over the PCIe input/output (I/O) bus using the NVMe protocol, plus the time required by the SSD itself to read data from the storage media.

Time spent by the operating system isn't just time waiting, it's also time in which the CPU is busy. Imagine working on a system with a 3 GHz CPU. In this example, 1 microsecond is equivalent to 3,000 clocks. That means 10 microseconds of software latency equals 30,000 clocks spent not executing the user application. No matter how fast the hardware is, that software overhead is about the same. To get around that overhead, a new approach is needed.

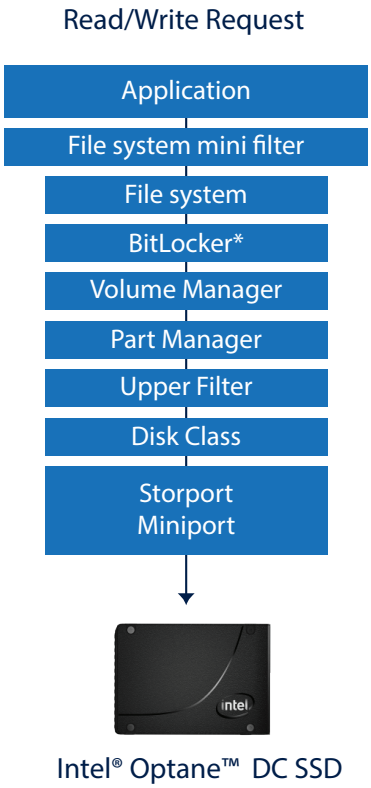


Figure 2. The storage stack is a significant part of Intel® Optane™ SSD latency

Bypassing Software Overhead with Intel® Optane™ DC Persistent Memory

By deploying Intel Optane technology as persistent memory instead of an SSD, much of that software overhead can be eliminated (see Figure 3). Intel® Optane™ DC persistent memory directly attaches to the memory channel instead of the PCIe bus. Intel Optane DC persistent memory is a new tier of memory that sits on the faster memory bus, in a module form factor, but has some qualities of storage (persistence).

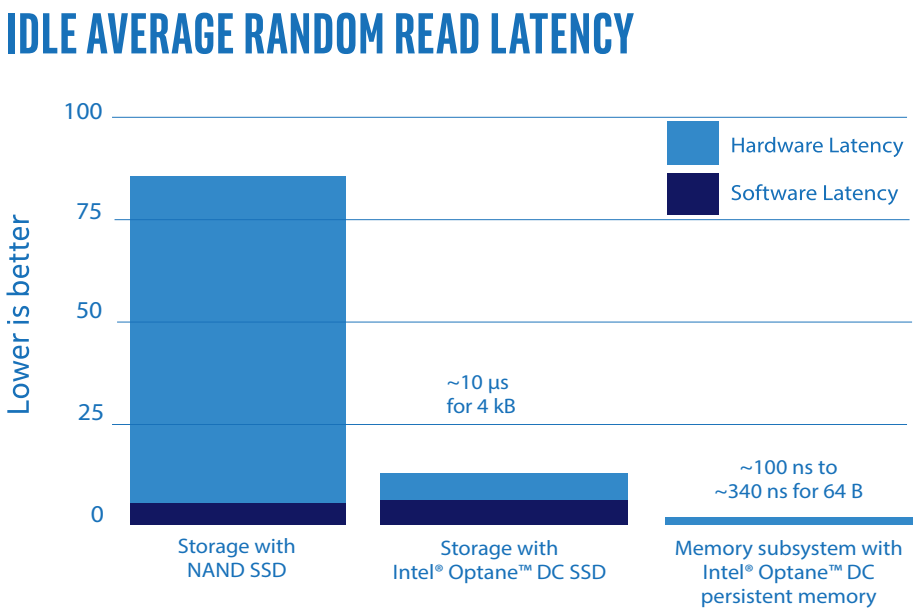


Figure 3. Comparison of read latency for NAND SSDs, Intel® Optane™ DC SSDs, and Intel Optane DC persistent memory.[2,3]

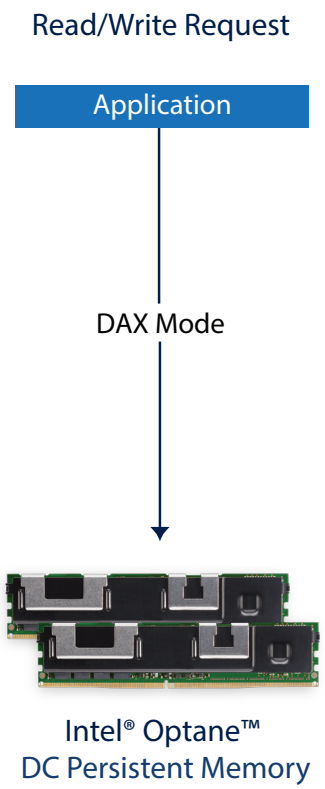


Figure 4. Intel® Optane™ DC persistent memory eliminates most software latency

Intel Optane DC persistent memory avoids software overhead by enabling mapping of persistent data for access directly from the CPU with memory load and store instructions—bypassing the operating system storage stack when reading and writing data. The operating system is directly involved in setting up access, providing a pointer to the location of the data, but need not act for regular reads and writes, resulting in much faster, direct load/store access to large capacity byte addressable space for working data.

The highly optimized architecture of Intel Optane DC persistent memory module reduces latency considerably. This module contains a specially designed memory controller with a highly optimized hardware only read/write path and the ability to connect to the customer memory controller in the CPU itself.

With the new media and optimized design, idle average latency of less than 100 ns if the data is found in DRAM, and up to 340 ns if it is found in Intel Optane DC persistent memory (in Application Direct Mode).[1] The latency is specified as a range because these two memory types work together as a hierarchy (see Figure 4).

In addition, because persistent memory is treated by the system as memory rather than storage, data requests are made with smaller, more efficient 64-byte accesses. Each load or store instruction fetches a 64-byte cache line with low latency. At the Intel Optane DC persistent memory module, this results in a read or write of 256 bytes of data accessed—much smaller than the 4 KB accessed in an Intel Optane DC SSD.[3] This means you can access data with smaller granularity and less overhead, and software developers can work with smaller structures to streamline their applications.

Intel® Optane™ Technology Performance for High-Throughput Uses

While the average-idle-latency advantages of Intel Optane technology are impressive, real applications access data at high rates with concurrent reads and writes. In the presence of this type of heavy traffic, these applications especially appreciate low-latency accesses. To understand the performance advantage of Intel Optane DC persistent memory, we return to the latency at load plots used to analyze the differences between SSDs. To simplify the discussion, we'll focus now only on the average latencies. Figure 5 shows just such latency for different system throughputs to a single SSD or memory module. The three technologies pictured, with latencies progressively lower by 10x or more at any throughput, offer a compelling storage hierarchy upon which applications can excel.

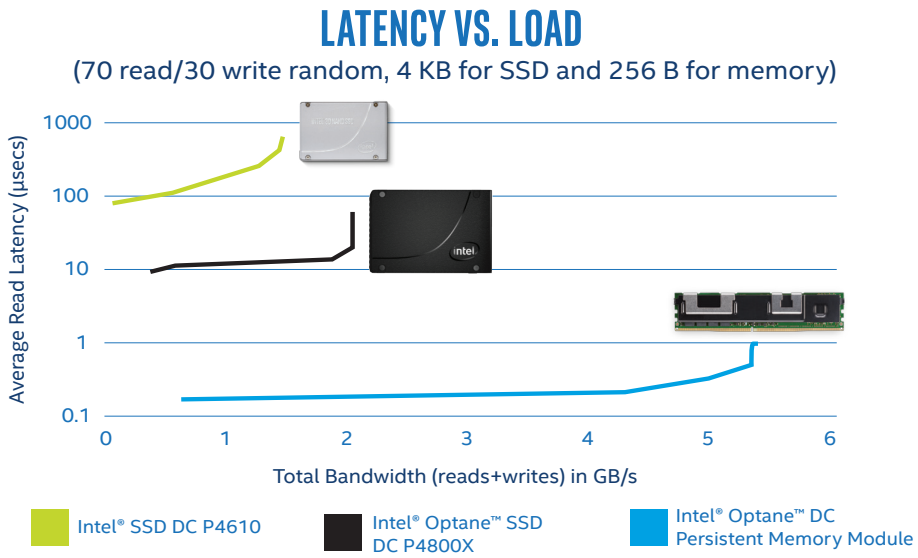


Figure 5. Comparison of the loaded latency for Intel® NAND SSDs, Intel® Optane™ DC SSDs, and Intel Optane DC persistent memory.[1, 3]

Intel Optane DC persistent memory offers by far the lowest latencies at all bandwidths for a truly unique performance with persistence.[1, 3] The very low average latency (blue line) offered by the Intel Optane DC persistent memory represents a truly new class of storage. This latency was already stated as between 100ns and 340ns depending on access locality. From the chart it's clear that the latency measured falls within the lower part of that range. That's because we measured 256 byte accesses made up of four cache lines, so while one cache line was amiss, the other three are hits, for an average latency at the lower fraction of the range. Notice that this low latency is consistent as the load is increased substantially. Even a single memory module can sustain throughputs significantly higher than either SSD tested.

Putting Persistent Memory to Use in the Data Center

There are many practical applications for Intel Optane DC persistent memory because it offers consistent low-latency, non-volatility, and high capacities. That makes it ideal for in-memory database applications that need mission-critical performance for large data volumes. For example, to remain competitive, financial services institutions rely on constantly changing trading data correlated with massive amounts of historical data. To process information quickly, these businesses frequently employ in-memory databases, using DRAM to provide real-time insights. But using DRAM has several drawbacks, including:

- Higher per-gigabyte costs compared to Intel Optane DC persistent memory
- Limited capacity per DIMM
- Volatility, which means the massive dataset must be loaded into memory after every restart or power failure before the system is again available.

In contrast, Intel Optane DC persistent memory offers:

- Lower cost per gigabyte, enabling an expanded data size within the same budget, or depending on the application a similar performance at lower cost.
- Much higher capacity memory, with up to 512 GB DRAM per module
- Persistence for rapid recovery times after a power outage or system reset

With the larger capacity it makes available, Intel Optane DC persistent memory can dramatically change the data-tiering landscape used by financial-services institutions and large enterprise businesses. Instead of migrating warm data from less costly NAND SSDs into DRAM as needed, organizations can create a large, non-volatile hot data tier in memory, giving data scientists and analysts real-time or near-real-time access to critical insights over that larger data set.

Conclusion

Intel Optane technology is the first new memory technology to be produced in volume in decades. It has entered the platform as a solid state drive, accessible through the standard OS storage stack and APIs. As an Intel Optane™ SSD, it delivers latency an order of magnitude lower than NAND SSDs and an even greater quality of service advantage. This new memory technology is also now available as system memory as part of an optimized system paired with 2nd Generation Intel® Xeon® Scalable processors. It is accessible directly from the applications without operating system overhead. Intel Optane DC persistent memory delivers bigger capacity than DRAM and persistence at extremely low latencies. These technologies, with latencies progressively lower by 10x or more at any throughput, offer a compelling storage hierarchy upon which applications can excel. Together, NAND SSDs, Intel Optane SSDs and Intel Optane Persistent memory fill out the memory and storage hierarchy enabling data to be placed at the right level so that it may be accessed quickly to deliver excellent system performance.

Intel Fellow Frank Hady

Frank Hady is an Intel Fellow and the Chief Optane Systems Architect in Intel's Non-Volatile Memory Solutions Group (NSG). Frank leads research and definition of Intel® Optane™ technology products and their integration into the computing system. Frank has served as Intel's lead platform I/O architect, delivered research foundational to Intel® QuickAssist Technology, and driven significant platform performance advances. He has authored or co-authored more than 30 published papers on topics related to networking, storage, and I/O innovation and presents often on memory and storage. He holds more than 30 U.S. patents. Frank received his bachelor's and master's degrees in electrical engineering from the University of Virginia, and his Ph.D. in electrical engineering from the University of Maryland.

Learn More

Learn more about how Intel Optane technology is disrupting the memory and storage hierarchy in the data center by exploring other papers in the Memory and Storage Technical Series.

To learn more about Intel Optane DC persistent memory, visit: **https://www.intel.com/content/www/us/en/products/memory-storage/optane-dc-persistent-memory.html**

To learn more about Intel Optane DC SSDs, visit: **intel.com/content/www/us/en/products/memory-storage/solid-state-drives/data-center-ssds/optane-dc-ssd-series.html**

Additional reading – Intel® Optane™ Technology: Memory or Storage? Both.

1.Source – Performance results are based on Intel testing as of November 15, 2018. Measured using FIO 3.1. Common Configuration - Intel 2U Server System, OS CentOS 7.5, kernel 4.17.6-1.el7.x86_64, CPU 2 x Intel® Xeon® 6154 Gold @ 3.0GHz (18 cores), RAM 256GB DDR4 @ 2666MHz. Configuration – Intel® Optane™ SSD DC P4800X 375GB and Intel® SSD DC P4610 3.2TB. Intel Microcode: 0x2000043; System BIOS: 00.01.0013; ME Firmware: 04.00.04.294; BMC Firmware: 1.43.91f76955; FRUSDR: 1.43. The benchmark results may need to be revised as additional testing is conducted.

2. Source – Performance results are based on Intel testing as of July 24, 2018. Average read latency measured at queue depth 1 during 4k random write workload. Measured using FIO 3.1. comparing Intel Reference platform with Intel® Optane™ SSD DC P4800X 375GB and Intel® SSD DC P4600 1.6TB compared to SSDs commercially available as of July 1, 2018.

3. Source – Performance results are based on Intel testing on February 20, 2019. Configuration: Intel® C620 Series Chipset, 28-core Intel® Xeon® Scalable processor (QDF QQYZ), 2,666 megatransfers per second (MT/s), 256 GB, 18 W, 32 GB DDR4 DRAM (per socket), 128 GB Intel® Optane™ DC persistent memory (per socket), firmware: 5336, BIOS: 573.D10, WW08 BKC, running Linux OS 4.20.4-200.fc29. Performance tuning quality of service (QoS) disabled, IODC=5(AD).

Performance results are based on testing as of the date set forth in the configurations and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit **intel.com/benchmarks**.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No product or component can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

 Printed in USA 1219/JR/RA/PDF Please Recycle 338865-003US